Evidence for adult age-invariance in associative false recognition

Article (Accepted Version)

Pansuwan, T, Breuer, F, Gazder, T, Lau, Z, Cueva, S, Swanson, L, Taylor, M, Wilson, M and Morcom, A M (2019) Evidence for adult age-invariance in associative false recognition. Memory. pp. 1-15. ISSN 0965-8211

Evidence for adult age-invariance in associative false recognition

Pansuwan, T[1]., Breuer, F[1]., Gazder, T[1]., Lau, Z[1]., Cueva, S[1]., Swanson, L[1]., Taylor, M[1]., Wilson, M[1]., Morcom, A.M[2].

[1]Psychology, University of Edinburgh.

[2]School of Psychology, University of Sussex.

Author Note

The first four authors contributed equally to the paper. Correspondence regarding this article should be addressed to Alexa M. Morcom, School of Psychology, University of Sussex, Pevensey 1, Brighton BN1 9RH, United Kingdom. Email: a.m.morcom@sussex.ac.uk

Word count = 9822 (excluding abstract, tables & references)

Keywords

False memory, recognition, semantic, ageing, DRM

Abbreviations

Deese-Roediger-McDermott (DRM), Mean (M), Standard deviation (SD), Bayes Factor (BF)

Abstract

Older people are more prone to memory distortions and errors than young people, but do not always show greater false recognition in the Deese-Roediger-McDermott (DRM) task. We report two preregistered experiments investigating whether recent findings of age-invariant false recognition extend to designs in which studied items are blocked. According to Tun, Wingfield, Rosen, & Blanchard (1998), age effects on false recognition in the DRM task are due to a greater reliance on gist processing which is enhanced under blocked study conditions. Experiment 1 assessed false recognition in an online variant of the DRM task where words were presented visually, with incidental encoding. The results showed Bayesian evidence against greater false recognition by older adults, whether lures were semantically associated with studied lists, or perceptually related (presented in the same distinctive font as studied lists) or both. Experiment 2 used a typical DRM procedure with auditory lists and intentional encoding, closely reproducing Tun et al.'s (1998) Experiment 2 but omitting an initial test of recall. The results showed evidence against an age-related increase in critical lure false recognition under these conditions. Together, the data suggest that older people do not make more associative memory errors in recognition tests than young people.

Introduction

It is well known that memory for events declines in later life, so there is keen interest in understanding aspects which are preserved. An altered interplay between memory and accumulated knowledge in older people has been proposed to accompany declining specific memory. While knowledge may support preserved memory for the gist (general meaning) of events (Castel, 2005; Naveh-Benjamin, Craik, & Guez, 2005; Reder, Paynter, Diana, Ngiam, & Dickison, 2007; Umanath & Marsh, 2014), reliance on gist brings a cost in terms of increased memory distortions (for review see Devitt & Schacter, 2016; McCabe, Roediger, Mcdaniel, & Balota, 2009; Stark & Squire, 2000). In support of this view, older people are more likely to falsely recognise images of objects when they have previously encountered other objects from the same semantic category, for example a different dog when several dogs were studied (Koutstaal & Schacter, 1997; Koutstaal, Schacter, Galluccio, & Stofer, 1999). They are also typically reported as being more prone to errors in the Deese-Roediger-McDermott (DRM) word list task (Deese, 1959; Roediger & McDermott, 1995), in which people falsely remember critical lures which are characterised by a strong associative relation to studied items, for example doctor, when a list including hospital and needle was studied (Intons-Peterson, Rocchi, West, McLellan, & Hackney, 1999; Norman & Schacter, 1997; Schacter, Israel, & Racine, 1999; Tun, Wingfield, Rosen, & Blanchard, 1998; see Gallo, 2006 for review).

In this report we address a challenge to this view: the proposal that associative false recognition may be age-invariant. Gallo (2006, pp.182-188) noted that the common practice of testing recognition after recall introduces a potential confound for measures of age effects on false recognition. Direct comparisons in young adults suggest a testing effect for false as well as true recognition (Gallo, 2006, pp. 147-150; Huff, Coane, Hutchison, Grasser, & Blais, 2012; Roediger & McDermott, 1995). This is potentially important as it would mean that if false recall were increased in older people on an intervening test, their subsequent false recognition would also be increased. Gallo's (2006) review found that age-related differences were small or absent in 14 published experimental conditions which tested recognition only, with a 3% mean increase in false recognition in older relative to young adults (whether assessed on raw or on adjusted proportions, i.e., correcting for response bias by subtracting proportions of false alarms to novel items). These age effects were not significant on unweighted tests of individual study means, unlike the 7% difference in the studies that tested recall first. Although the number of available datapoints was low, findings in DRM conditions without prior recall remain mixed in more recent studies. For example in an investigation of the effects of item repetition, Skinner & Fernandes (2009) observed the predicted age-related increase in false recognition when items were studied 3 times, but no difference in the control, single-repetition condition. In contrast, Wong & Gallo (2016) found a reliable age effect across two experimental conditions.

These results from DRM studies appear to differ from those of categorised pictures studies. In all cases the latter have (to our knowledge) found age-related increases in false recognition (see (Pidgeon & Morcom, 2014, for detailed review). When unstudied lures are novel exemplars of the same semantic categories as studied items, and sets of 9 per category have been studied (compared to usually 10 or more in DRM tasks), older adults typically falsely recognise 15-20% more than the young (e.g. Koutstaal & Schacter, 1997; Koutstaal et al., 1999; Pidgeon & Morcom, 2014). Moreover, this age effect is greater for pictures of familiar objects than for pictures of unfamiliar, novel objects which are perceptually similar but for which minimal prior knowledge is available. This supports the notion of specifically enhanced semantic gist processing, over and above perceptual errors, in older people (Koutstaal et al., 2003; Pidgeon & Morcom, 2014). However, an increase in false recognition of abstract lures when multiple perceptually related items had been studied also pointed to a possible sensitivity to perceptual errors (Pidgeon & Morcom, 2014).

The contrast we draw between DRM and categorised pictures tasks and materials is superficially at odds with the findings of a large recent meta-analysis of age effects on false recognition over 232 experiments (Fraundorf et al., 2019). The authors found larger age-related differences across studies that used semantically-related rather than unrelated lures, and no evidence for a specific role of surface level, perceptual similarity. These are important results but are difficult to compare with many of the findings in the literature because the measure used in the meta-analysis was discrimination (d') of studied items versus lures. This means that the results of the meta-analysis reflect combined age effects on true and false recognition.

We were initially interested in whether an age-related increase in false recognition would be found in a DRM task using design and procedures like the categorised picture studies (Burnside, Hope, Gill, & Morcom, 2017). Burnside et al. (2017) adapted a paradigm from Arndt & Reder (2003) to measure associative and perceptual false recognition using visually (rather than auditorily) presented words written in elaborate, distinctive fonts. The items in the word lists associated with each critical lure were also intermixed rather than presented consecutively (blocked) during encoding. Lastly, a specific orienting task at study ensured processing of perceptual and semantic aspects of the stimuli: in this case by judging how well words' meanings 'fit' their font. These three procedural elements differed from a typical DRM task. Burnside et al. found Bayesian evidence against increased false recognition of critical lures in older people, who instead were more likely to falsely recognise the Font-only lures which were presented in the same font but not semantically related to studied items. Thus, they found that in this task age-related memory errors were driven more by perceptual than semantic relatedness. The results converge with those of another study that used intermixed visual word lists without distinctive fonts: Wilson, Potter, & Cowell (2018) also found no significant age-related increase in false recognition of critical DRM lures, although memory was measured solely in terms of discrimination between studied items and lures (as in Fraundorf, Hourihan, Peters, & Benjamin, 2019).

The available data suggest that age effects on associative false recognition are less prominent than age effects on false recognition of categorically-related familiar pictures, at least under comparable conditions. This contrast has implications for theories of false memory. The two main accounts, Fuzzy Trace Theory and Activation-Monitoring Theory, explain semantically-related lure errors in terms of gist memory traces or associative activation, respectively. These act in opposition to verbatim memory traces (in Fuzzy Trace Theory) and monitoring processes (in both; Reyna & Brainerd, 1995; Roediger & McDermott, 1995). It remains in question whether semantic gist or associative representation is more fundamental to memory illusions (e.g., Brainerd, Yang, Reyna, Howe, & Mills, 2008; Cann, McRae, & Katz, 2011; Huff & Hutchison, 2011; Hutchison & Balota, 2005; Roediger et al., 2001). If age-related recognition errors depend on whether the relation between the studied item set and the lure is predominantly semantic or only associative, this would support the idea that multiple kinds of relation contribute (Carneiro, Garcia-Marques, Lapa, & Fernandez, 2017; Coane, McBride, Termonen, & Cutting, 2015; Montefinese, Zannino, & Ambrosini, 2014), contrary to Activation Monitoring Theory's fully associative account. However, such a result would suggest that the idea of semantic gist in memory requires additional specification (Brainerd et al., 2008).

In terms of theories of cognitive ageing, a finding that the age-related increase in false recognition does not extend to the DRM task would contradict the idea outlined above that older people rely more on prior knowledge. Reliance on prior knowledge also cannot explain errors due to visual perceptual similarity (Burnside et al., 2017; Pidgeon & Morcom, 2014) or phonological similarity (Watson, Balota, & Sergent-Marshall, 2001; D. M. Wilson et al., 2018). An alternative view is that older people's memory representations are in some way less specific (Abdulrahman, Fletcher, Bullmore, & Morcom, 2014; Burke & Barnes, 2018; Reagh et al., 2018; Trelle, Henson, Green, & Simons, 2017; D. M. Wilson et al., 2018; I. A. Wilson, Gallagher, Eichenbaum, & Tanila, 2006; Yassa, Mattfeld, Stark, & Stark, 2011). If this is correct, they may have difficulty discriminating between experiences with overlapping features, whether the features are semantic or perceptual. On this view, errors elicited by DRM critical lures may not increase, since these lures are dissimilar to many studied items (e.g. "hospital" and "needle" are strongly associated with the critical lure "doctor" but do not share many semantic features).

Despite these observations, it is not yet clear that false recognition in the DRM task is age-invariant, since other evidence points to the importance of key procedural variables. In an influential study, Tun, Wingfield, Rosen & Blanchard (1998) found an age effect only for blocked lists (Experiment 2), contrasting with a null finding for intermixed lists (Experiment 3). The results suggested that older people relied more on semantic gist than young people only when blocked encoding encouraged gist processing. Word lists were also presented auditorily with short stimulus onset asynchrony (SOA; 1.5 seconds). Auditory presentation is thought to limit opportunities for item-specific processing relative to relational (gist) processing (Smith & Hunt, 1998), and may differentially affect older and younger adults, as may the available time for encoding (Craik & Rabinowitz, 1985). A recent study suggested that the null effect of age on associative false recognition may extend to a blocked list procedure using visual presentation (Coane, Huff, & Hutchison, 2016), although evidence for the null hypothesis was not formally assessed. Tun et al.'s findings are also qualified by the fact that they assessed recognition after an initial recall test in which the older group recalled more critical lures (Gallo, 2006). Moreover, Tun et al. did not directly test age-related differences in novel-adjusted false recognition of critical lures, although there was a significant interaction between condition and age group across studied items and all lure types.

In the two experiments reported here, we aimed to close the gap between studies that have used procedures comparable to the categorised pictures task (e.g., Burnside et al., 2017; D. M. Wilson et al., 2018), and studies using a more typical DRM task (e.g., Kensinger & Schacter, 1999; McCabe & Smith, 2002; Norman & Schacter, 1997; Tun et al., 1998). Experiment 1 sought to clarify whether older people still show the same levels of associative false recognition as young people when lists are blocked. We retained visual presentation and incidental encoding like Burnside et al. (2017), but added a plain-font *DRM-only* condition, omitting the distinctive visual information provided by the elaborate fonts. This addition also allowed us to test whether the age effect on perceptual false recognition would generalise when the incidental task performed at encoding was changed to a pleasantness judgment suitable for plain-font as well as distinctive-font conditions (an orienting task also used in many categorised pictures studies, e.g. Koutstaal et al., 2003; Koutstaal & Schacter, 1997; Pidgeon & Morcom, 2014). We made two sets of alternative predictions. If Tun et al. (1998) were correct that older people are more prone to gist memory than young adults when word lists are blocked, an age effect on false recognition would now be observed, unlike in both Burnside et al.'s (2017) experiments and (D. M. Wilson et al., 2018). Alternatively, if associative false recognition is age-invariant when there is no prior recall test, we would find Bayesian evidence in favour of the null hypothesis, like Burnside et al. (2017). A third possibility was that older people do more readily rely on gist, but that list blocking is not the only relevant variable. In Experiment 2, we therefore reproduced Tun et al.'s (1998, Experiment 2) auditory lists design and procedure as closely as possible, using intentional encoding and a short SOA, but omitting the initial recall test. We predicted that if Tun et al.'s hypothesis is correct, we would replicate the age-related increase in false recognition they reported, despite the lack of a recall test. Otherwise, we would again find age-invariance.

Experiment 1

Methods

Preregistration

The study was preregistered on the Open Science Framework after most data were collected but prior to data processing (https://osf.io/gpu7b/).

Participants

Participants were 188 adults: 103 young (18-33 years, *M* = 21.9, *SD* = 2.38, 72 female) and 85 older (60-75 years, *M* = 67.3, *SD* = 4.09, 57 female). Following data exclusions (see Data analysis), there were 34 young and 33 older participants in the DRM condition, 38 and 23 in the Font condition, and 31 and 29 in the DRM-font condition. Across conditions, the distribution of levels of education differed according to age ($\chi^2(4) = 29.2$, $p < 0.001$). Unlike in many laboratory studies, young people overall had had *more* education: the modal level was an undergraduate degree in both groups, but the spread was greater in the older group. The young were much more likely to have undergraduate degrees (87 compared to 41 older people), less likely to have non-degree level (1 versus 17) or Master's level education (10 versus 15) (see Supplemental Material for exploratory analyses relating to education level). Participants gave informed consent at the start of the initial questionnaire and were not paid. The study link was distributed via University channels, personal contacts and social media. The study was approved by the PPLS Research Ethics Committee, University of Edinburgh (ref. 33-1718/4).

We used a time-based stopping rule for data acquisition. Sensitivity analysis was done in (G*Power 3.1.9.2; Faul, Erdfelder, Lang, Buchner, & Kiel, 2007), with alpha = .05. Given the N, we had .8 sensitivity to an effect of size f = .228 for the interaction of age group x condition. With 65 young and 62 older participants across the 2 DRM conditions, we also had sensitivity of .8 with a directional (one-tailed) effect for d = .443, i.e. a moderate-sized age-related increase in associative false recognition across the two DRM conditions. For the Font condition we had .8 (one-tailed) sensitivity to detect a medium to large sized age-related increase of d = .664.

Materials

Questionnaires. Participants first indicated their gender, age, and native language, then responded to brief questions about health and medication (see Data analysis for questions and exclusion criteria). They also reported their education level (1 = GCSE, Scottish Highers or O Level, 2 = A/AS or Scottish Advanced Highers, 3 = Undergraduate degree, 4 = Master's degree, 5 = PhD or doctorate, 0 = None of the above). After completing the experimental task, non-native English speakers completed an additional 5-part question on their current command of English taken from a Language Questionnaire by Vega-Mendoza, West, Sorace, & Bak (2015), rating their Expression, Comprehension, Reading and Writing on a 6-point scale from 0 = None to 5 = Fluent; see Data analysis for inclusion criterion).

Experimental task. Stimuli were visually presented words written either in a distinctive font or a standard font (Arial). Words were taken from 128 thirteen-word associated DRM lists taken from the University of South Florida (USF) Free Association Norms database (Nelson, McEvoy, & Schreiber, 2004); http://w3.usf.edu/FreeAssociation/). We selected lists with the highest mean backward associative strength (BAS) from their top 13 items to the critical lures. The mean BAS across all lists was .323 (*SD* = .092). Lists were also screened for Americanisms, offensive words and variations of the same word. Twenty-four Google fonts (from http://fonts.google.com/) were chosen to have maximal distinctiveness and readability.

There were three experimental conditions. In the DRM and Font-DRM conditions, sets of associated words from DRM lists were studied. In the DRM condition, all words were studied in Arial, while in the Font-DRM condition, each set was studied in the same distinctive font (different for each set). In the Font-only condition, sets of unrelated words were studied in the same distinctive font. Test lists included studied items, unstudied items related to studied sets (lures) and unrelated unstudied (novel) items. In the DRM condition all test items were presented in Arial. In the Font-DRM condition, lures were associated with and shown in the same font as their studied set. In the Font-only condition, lure words were drawn from unstudied DRM lists and shown in studied fonts, so were only perceptually related to studied items. In all conditions the novel items were not associated with the studied items and except in the DRM condition they were shown in unstudied distinctive fonts.

The experiment consisted of 2 study-test cycles. In each cycle, different complete DRM and/or Font item sets were studied, and their corresponding items tested (48 items at study and 12 at test). Stimulus lists were constructed so that allocation of items to critical lure and matched novel item conditions, and allocation of items to studied and matched novel-item conditions, was fully counterbalanced. Twenty-four of the 128 DRM lists were randomly selected for each set of study and test lists in the DRM and Font-DRM conditions. Study lists comprised 8 twelve-word associated DRM lists (total = 96; list positions 1 to 13, excluding 6). Test lists comprised 3 studied items from each of the 8 studied lists (list positions 1, 8 and 10; total = 24 items), 8 critical lures corresponding to each studied list, 8 weakly associated (‘weak’) lures which were unstudied members of the 8 studied DRM lists (list position 6), and 16 novel items (total = 56). Half the novel items were critical-matched (critical lures from unstudied DRM lists) and half were list-matched (list position 6 items from unstudied DRM lists).

Each set of Font-only study and test lists was constructed from all 128 DRM lists. The study lists comprised 8 unrelated 12-word sets (total = 96). Sets were constructed by taking one word from each of 12 DRM lists (using items from the same list positions as in the DRM conditions). Test lists comprised 3 studied items per set, with 16 critical and 16 weak lures and 16 critical-matched and list-matched unrelated novel items taken from the remaining 32 DRM lists (the same list positions were used as for the DRM conditions).

DRM lists and fonts were randomly allocated to conditions 8 times to create 6 counterbalanced stimulus versions for each of the 3 conditions. Consecutive participants were allocated randomly to a condition and list version.

Procedure

The experiment was conducted online in a browser on a Mac or PC computer with a keyboard, programmed in Javascript (version 5.0.3). There were two study-test cycles. At study, participants judged the pleasantness of each word shown on-screen at a time using a 3-way button press ('Unpleasant', 'Neutral' and 'Pleasant'). During a filler task, they then completed two arithmetic 4-choice questions. At test, they completed a recognition memory task using a variant on the Remember-Know procedure (Gardiner & Java, 1991; Migo, Mayes, & Montaldi, 2012); see https://osf.io/r5nqu/ for instructions (we do not report data separated by recollection and familiarity reported here, see Data analysis). During study and test, items remained on-screen for up to 7000 ms or until a response was provided. Each response was followed by appearance of a central crosshair which stayed on-screen for 1500 ms. Response keys were 'A', 'Z', 'M' and 'K' on the computer keyboard and allocations to response categories were counterbalanced over participants.

Data analysis

Pre-processing. The participant survey data (392 entries) were first checked manually and 270 valid entries found (156 young, 114 older). Seventy-two were initially discounted from further analysis. Of these, 9 were incomplete, and 63 were excluded on *a priori* criteria: 42 gave invalid ages (not between 18-33 or 60-75 years), 5 young and 13 older adult entries were excluded on medical criteria (diabetes, open heart surgery, epilepsy, brain surgery, in-hospital treatment for mental health problems, head injury with > 1 hour loss of consciousness, medication which may cause drowsiness), and 3 young adult entries were excluded due to a reported English fluency score of < 15/20. We then checked for duplicates and excluded a further 50 entries. Sixty-six survey entries shared an IP address with another entry. If demographic information was consistent with the same individual having completed the survey more than once, we checked for multiple sets of task data. Excluded survey entries had either multiple task data files or no data files. There were also 6 judged not to be duplicates, and of the duplicates, 10 yielded valid task datasets (9 where only first survey entries were associated with task data, 1 where words and fonts studied during a short first attempt could be excluded from analysis of a full second attempt).

Next, the task data files were cross-referenced with the valid survey entries via unique filename identifiers logged in the survey, yielding 198 participants with both task data and survey entries (112 young, 86 older; a further 72 survey entries were not accompanied by task data files). The task data were inspected, and 4 participants with empty data files excluded. *A priori* criteria were then applied to exclude participants with very poor memory overall (hits - unrelated novel false alarms (FA) < .1) (2 young and 1 older), or less than 5 valid trials (with responses) per condition with which to calculate proportions (3 young). Lastly, individual test trials were discounted from further analysis in cases where those trials' study phase was invalid (i.e., absent data entries or no logged response; 1-4 trials for 4 young and 4 older participants, or some items had been studied in an initial, first attempt study phase; 13 trials in 1 older participant, see above).

Statistical analysis. Data were extracted and analysed in R (version 3.5.0, https://www.r-project.org/) and JASP (version 0.9.1, https://jasp-stats.org/; (Rouder, Speckman, Sun, Morey, & Iverson, 2009). Analyses of variance (ANOVAs) used type III sums of squares. Greenhouse-Geisser corrections were applied to degrees of freedom and p values for *F*-tests where relevant, and Welch corrections were applied to *t*-tests. *A priori* null-hypothesis significance tests used alpha = .05, and exploratory tests used alpha = .005. All tests are 2-tailed unless otherwise specified. We also computed Bayes Factors (BF) to assess evidence for the null hypothesis and for key conclusions in favour of the alternative hypothesis. For clarity (although Bayesian evidence is continuous) we adopt the labels used in JASP (Lee & Wagenmakers, 2013): BF < 3 corresponds to “anecdotal” evidence, regarded as inconclusive, 3 < BF < 10 to “moderate”, 10 < BF < 30 to “strong”, 30 < BF < 100 to “very strong”, and BF > 100 to “extreme” evidence. Bayesian *t*-tests used uninformative Cauchly priors with M = 0, width = .71, with additional robustness checks of BFs under a range of prior distribution widths (width = .5 or .9) to assess whether this substantially changed the results (a wider prior increases evidence for the null hypothesis). The plots in Figures 1 and 2 were created using code from Allen, Poggiali, Whitaker, Marshall, & Kievit (2018).

The principal analyses of true and false memory accuracy used the signal detection theory (SDT) index d’ to assess the degree to which participants’ ‘old’ responses discriminated hits from unrelated novel items (true d’), critical lures from unrelated novel items (critical false d’), and weak lures from unrelated novel items (weak false d’) (Arndt, 2010). This approach is better than adjusting raw hits and false alarms by subtraction of unrelated novel false alarm proportions because it accounts formally for response bias without assuming a dual process model (Snodgrass & Corwin, 1988), and because d’ is normally distributed, unlike response proportions. For all d’ measures, hit and false alarm proportions were calculated from raw response frequencies after correction of zero scores by adding .5 to hits or false alarms, and 1 to total of studied or unstudied trials (Hautus, 1995; Snodgrass & Corwin, 1988). We ran an additional analysis of hit proportions for direct comparison with the results of Tun et al. (1998) and other studies which did not adjust for response bias (see Introduction). However, although we preregistered analyses of true and false recollection reflected in Remember responses, these were intended to assist interpretation of age-related differences in false recognition (see Burnside et al., 2017). Therefore, as we found no differences between groups in overall false recognition, we do not report analyses of recollection. Lastly, we conducted exploratory analyses of discrimination between studied items and critical lures, a measure which formed the basis of a recent meta-analysis (Fraundorf et al., 2019; see General Discussion).

Each d’ measure was computed from a proportion of either hits or lure false alarms and the proportion of false alarms to the corresponding (counterbalanced and matched) unrelated novel items. For True d’ these were hits and list-matched unrelated novel FA. For each experimental condition (DRM, DRM-font and Font) the main measure of false memory was Critical false d’, computed using critical lure and critical-matched FA (although in the Font condition the Critical lures were not associatively related to the studied lists, they were critical lures from unstudied lists; see Materials). In subsidiary analyses (see Supplemental Material) we also computed Weak false d’, using weak lure and list-matched unrelated novel FA.

For analyses of response times (RTs), we also performed a *z*-score correction to account for individual (and therefore also age-related) differences in processing speed that might impact the results (Faust, Balota, Spieler, & Ferraro, 1999). This is a concern particularly where age effects are predicted to be most pronounced in a condition with longer RTs such as Critical lure false alarms. For each participant, RTs on every trial were transformed by subtracting the global mean of that participant’s RTs across all trials, and dividing by the global standard deviation of their RTs across all trials. The results are reported in the Supplemental Material.

Results

Preregistered analysis

True and false memory accuracy proportions and response times (RTs) are summarised in Table 1, and Figure 1 illustrates age-related and condition effects for the d’ measures of critical lure false recognition.

[Table 1 near here]

False recognition: discrimination. The principal analysis evaluated Critical lure false recognition across the 3 conditions using d' (Fig. 1). Older people did not show greater false recognition of DRM critical lures or Font 'critical' lures (Fig X). Analysis of variance (ANOVA) with factors of Age (Young, Older) and Condition (DRM, DRM-font, Font) gave rise to a main effect of Condition ($F(2,183) = 41.3$, $p < 0.001$, $\eta^2_p = 0.311$) which reflected greater false recognition for the two DRM conditions than for the Font condition (for DRM vs. Font, Tukey's $t = 7.02$, $p < 0.001$, $d = 1.33$; for DRM-font vs. Font, t = 8.61, $p < 0.001$, $d = 1.67$; for DRM vs. DRM-font, t = 1.81, $p = 0.168$). Effects of Age were not significant (for main effect, $F(1,183) = .726$, $p = 0.395$; for interaction, $F(2,183) = .059$, $p = 0.943$). In the Bayesian ANOVA there was also evidence against inclusion of either the main effect of Age or the interaction in the model ($BF_{01} = 6.18$ and 11.1). A *t*-test across conditions also confirmed evidence in favour of a null effect ($BF_{01} = 6.32$; with prior width = .50, $BF_{01} = 4.61$). These results demonstrate strong associative and weaker perceptual false recognition effects which did not differ between age groups.

To maximise N to test for age effects on associative false recognition, we also assessed Critical lure d' across both DRM conditions. Again the groups did not differ significantly ($t(124.4) = .738$, $p = 0.462$), and there was moderate Bayesian evidence against an age-related increase ($BF_{01} = 8.52$; with prior width = .5, $BF_{01} = 6.20$).

[Figure 1 near here]

True recognition: discrimination. Analysis of True d' with factors of Age and Condition showed no significant effects (for Age, $F(1,183) = .974$, $p = 0.325$; for Condition, $F(2,183) = 2.42$, $p = 0.0918$; for interaction, $F(2,183) = 1.59$, $p = 0.206$). There was equivocal evidence against inclusion of Condition in the model ($BF_{01} = 2.73$), but moderate evidence against inclusion of Age (BF_{01} = 5.21) and strong evidence against an interaction (BF_{01} = 10.5). This showed that true recognition did not differ according to age.

True and false recognition: raw accuracy. We also tested the effect of age on raw false alarm proportions in the DRM condition alone. This replicated the approach taken by Tun et al. (1998, Experiment 2) apart from our use of two rather than one unrelated novel item conditions. A mixed ANOVA with factors of Age (Young, Older) and Item Type (Hits, Critical lure FA, Weak lure FA, Critical-matched novel FA, List-matched novel FA) revealed a main effect of Item Type ($F(2.3,148.8) = 454.8$, $p < 0.001$, $\eta^2_p = .875$), reflecting more frequent FA across the two age groups to Critical lures than the other unstudied item types (for Weak lures, $t = 10.1$, $d = 1.23$; for Critical-matched unrelated novel $t = 12.4$, $d = 1.51$, for List-matched unrelated novel $t = 11.3$, $d = 1.38$; all Bonferroni-corrected $p < 0.001$), and more frequent FA to Weak lures than Critical-matched and List-matched novel ($t = 5.37$, $d = .656$; $t = 4.34$, $d = .531$). The two novel item types did not differ significantly ($t = 1.85$, $p = 0.684$). Hits were also more frequent than all types of FA ($t > 15$, $p < 0.001$). The only age effect was a main effect ($F(1,65) = 7.55$, $p = 0.0078$, $\eta^2_p = .104$) reflecting a greater tendency to respond 'old' by the young group to studied and unstudied items (Table 1). The interaction of Age x Item type was not significant ($F(2.3, 148.8) = .967$, $p = 0.392$), and there was evidence against inclusion of this interaction in the model (BF_{01} = 4.11). These data converged with the discrimination analysis to show age-invariance in associative false recognition.

Exploratory analysis

Font condition false recognition. A separate comparison of Font lure false d' between the two age groups confirmed the impression from the overall ANOVA (see above) that there was no age-related increase in perceptually-driven false recognition (BF_{01} = 5.73 against the predicted increase; BF_{01} = 4.27 with prior width .5).

Discrimination between studied items and lures. Comparison between groups of d' for studied items versus critical lures showed evidence against decreased discrimination in older adults both for associative conditions (DRM and DRM-font; $BF_{01} = 4.33$; $BF_{01} = 3.23$ with prior width .5) but only anecdotal evidence for this null hypothesis in the Font-only condition ($BF_{01} = 2.14$).

True and false recognition: response times. Following Tun et al. (1998) we evaluated RTs for "old" responses, i.e. to hits and false alarms in conditions where there were enough trials (Table 1). Given the maximum of 8 trials per condition in lure conditions, in order to avoid bias by inclusion of only participants with high false recognition we did not apply any minimum number of trials for RT analysis. There were sufficient participants with RTs to studied and Critical lure item types for the DRM (33 young and 29 older) and DRM-font (30 young and 28 older) conditions but not for the Font condition, in which some participants made no 'Critical' lure false alarms (RTs available for 22 young and 13 older). ANOVA on median RTs with factors of Age, Condition (DRM, DRM-font) and Item type (Hit, Critical lure FA) revealed only significant main effects of Age Group ($F(1,116) = 17.8$, $p < .001$) and Item type ($F(1,116) = 81.6$, $p < .001$): RTs across conditions and item types were slower in older adults, and RTs across age groups and conditions were slower for Critical lures FA than for hits (for Condition, $F(1,116) = 1.656$, $p = 0.201$), but Age did not interact with Item type or with Item Type and Condition (for Age x Item type, $F(1,116) = .408$, $p = .524$, $BF_{01} = 1.54$, for Age x Condition, $F(1,116) = 1.23$, $p = 0.270$, $BF_{01} = 4.31$; for Condition x Item type, $F(1,116) = 1.47$, $p = 0.227$, for 3-way interaction, $F(1,116) = 2.08$, $p = 0.152$, $BF_{01} = 24.4$). Analysis of *z*-scored median RTs (see Supplemental Material) showed broadly the same pattern of findings, apart from (as expected) an absent main effect of Age. However, in that analysis the evidence against an interaction of Age and Item type was robust ($BF_{01} = 14.9$).

Discussion

In this experiment, there was consistent evidence against age-related increases in associative false recognition across the two DRM conditions. This is in line with Burnside et al.'s (2017) and Wilson et al.'s (2018) findings and Gallo's (2006) observations. The results further suggest that age-invariance in associative false recognition extends to situations where according to Tun et al. (1998) gist processing is increased in older adults, i.e. when related lists are presented together in blocks at study rather than intermixed (see also Coane et al., 2016). Moreover, the findings generalised to word lists presented in a simple standard font, as used in other DRM studies, rather than distinctive fonts. Although the first experiment was conducted online, overall levels of true and false recognition were consistent with those in Burnside et al.'s (2017) laboratory-based study: critical associative lure false recognition was around .5 (unadjusted; see Table 1 and Table 1 in Burnside et al., 2017), substantially higher than font lure false recognition (.2 unadjusted). With no instruction to make speeded responses, RTs were relatively slow (means around 1-2 sec for true recognition and 2-4 sec for critical FA; Table 1), but with a consistent pattern over trial types did not suggest inattention to the task. As in previous studies there was a clear delay in processing associative lures compared to studied items, but this effect did not differ significantly according to age, unlike in Tun et al.'s (1998) study. While evidence against a difference was equivocal for raw RTs, it was strong when RTs were scaled to remove individual differences in overall response times and therefore also effects due to age-related slowing.

The results for perceptual false recognition diverged from those of Burnside et al. (2017). Young and older people were equally likely to falsely recognise the visually similar lures. To our knowledge, no other studies have examined visually-related false recognition with verbal material in older people. With pictorial material, findings regarding perceptually-driven false recognition have differed, and may depend on how many similar exemplars have been studied (Pidgeon & Morcom, 2014). Previous studies have also reported age-related increases in false recognition of phonologically related lures in a DRM-type word list paradigm (e.g., Budson, Sullivan, Daffner, & Schacter, 2003). Several differences between the current study and Burnside et al.'s experiments may account for the discrepancy. Probably the most important is that by using a pleasantness task in the study phase instead of the font-fit task used by Burnside et al. (2017) – a change made to accommodate the standard font condition – participants were not required to attend to the distinctive fonts in the other two conditions. Other data suggest that older adults are less likely to process context spontaneously than the young (e.g. Craik, 1983; Taconnat, Clarys, Vanneste, & Isingrini, 2006). Thus, age-related increases in perceptual false recognition may have been paradoxically masked. Future studies can test this possibility with manipulations of the encoding task. The current study also varied condition between groups, not within participants, which might have interacted with the way that people processed the fonts.

While the determinants of perceptual false recognition in ageing require clarification, the results of Experiment 1 converged with those of Burnside et al. (2017) and Wilson et al. (2018) to support the suggestion that older people may not show increased associative false recognition, even with blocked list presentation. However, as this finding contrasts with data from some prior studies which differed procedurally from Experiment 1, it was essential to try to reproduce it using a standard DRM procedure. This was the aim of Experiment 2.

Experiment 2

Introduction

The typical DRM procedure emphasises gist processing in several ways which may differentially affect young and older people. Items from associated word lists are not only blocked together at study, but memory for each list may be tested separately. As noted in the Introduction, relatively fast auditory presentation may impede item-specific processing and emphasise age-related differences in encoding (Craik & Rabinowitz, 1985; Smith & Hunt, 1998). In addition, most DRM studies use intentional rather than incidental encoding, unlike the categorised pictures studies, recent DRM studies finding age-invariant false recognition (Burnside et al., 2017; Coane et al., 2016; D. M. Wilson et al., 2018), and the present Experiment 1. An item-specific orienting task such as a pleasantness judgement appears to reduce relational processing compared to an intentional task in which processing is unconstrained by instructions (Smith & Hunt, 1998), and age effects on true memory are often greater under intentional instructions (Old & Naveh-Benjamin, 2008; Spencer & Raz, 1995). There may be an accompanying increase in age effects on false memory, although available data suggest that this is not always the case (e.g. Thomas & Sommers, 2005),. To address these issues, we asked whether false recognition would be greater in older than young adults in a typical DRM procedure with these features. Again, we were interested in recognition performance in the absence of a confounding prior recall test. Otherwise, we based our procedures as closely as possible on those of Tun et al. (1998; Experiment 2), who observed an age-related increase in false recognition only for blocked study lists.

Methods

Preregistration

The study was preregistered on the Open Science Framework after 8 of 64 datasets had been collected and before any processing had been done (https://osf.io/vxym2/).

Participants

Participants were 31 young adults (18 - 23 years; *M* = 20.9, *SD* = 1.2; 24 female) and 32 older adults (64 - 80 years; *M* = 70.6, *SD* = 4.6; 21 female). Data from a further 2 young (1 technical failure, 1 poor performance) and 4 older adults (1 technical failure, 1 poor performance, 2 with insufficient valid trials) were discounted from analysis (see Preprocessing). Statistical power was estimated by sensitivity analysis because Tun et al.'s (1998) paper only reported the overall interaction of age group x condition (d = 1.1) for "old" responses across hits and false alarms. This effect may be larger than for false memory alone. With our intended samples of N of 32 per group (20 in the original study), we would have .8 power to detect a moderate effect (d = 0.36) for a group x condition test comparing critical lures and unrelated novel items only (alpha = .05; G*Power 3.1.7). The young participant with low performance was excluded after data collection finished, so final Ns were 31 and 32.

All participants were very fluent English speakers (5 younger adults were non-native). Young participants were recruited by local advertisement and a course credit system. Psychology students who had been taught about the DRM paradigm were excluded. Older participants were recruited via the department's Volunteer Panel and the University of the Third Age. Participants received no payment. All were in good self-reported physical and mental health, and not on any medication which could impair concentration. Ethical approval was obtained from the University of Edinburgh PPLS Research Ethics Committee (reference number: 179 – 1718/2).

The older group had had more years of formal education (since the first year of primary school) than the young (M = 18.1, SD = 2.18; for young, M = 15.7, SD = 1.62; $t(57.3) = 5.01$, $p < 0.001$, d = 1.26). They also differed in levels of education measured on a 6-point scale similar to that used for Experiment 1 (1 = GCSE, Scottish Standard or O Level, 2 = A/AS or Scottish Highers, 3 = Diploma such as DipCE, 4 = Undergraduate degree, 5 = Master's degree, 6 = PhD or doctorate, 0 = None of the above; young mode = 4, older mode = 5; χ^2 (3) = 21.3, $p < 0.001$). The groups were also compared on vocabulary as an estimate of crystallised intelligence, using the Wechsler Test of Adult Reading (WTAR; (Holdnack, 2001). This test updated the older WAIS-R Vocabulary Scale used by Tun et al. (1998), and provides age-adjusted norms to accommodate the usual increase in vocabulary from early to mid-life. The older group scored higher than the young on both raw performance (*M* = 46.9, *SD* = 1.93; for young, *M* = 40.7, *SD* = 4.87; $t(39.0) = 6.57$, $p < 0.001$) and age-corrected estimated verbal IQ (*M* = 120, *SD* = 3.69; for young, *M* = 112, *SD* = 8.72 for older, $t(40.1) = 5.01$, $p < 0.001$). (See Supplemental Materials for exploratory analyses of memory performance in relation to education level).

Materials

The ten associated (DRM) word lists each comprised 15 list items and one critical lure. Tun et al. (1998) drew words from Russell and Jenkins (1954), but as the items used were not published we used the more recent Stadler, Roediger and McDermott (1999) lists. The weakly associated lures used by Tun et al. (1998) were ranked 13 to 25 in the Russell and Jenkins (1954) word association norms. In the Stadler, Roediger and McDermott (1999) norms only items 13 to 15 were available, so these were used for all lists. The lists corresponded to the critical lures Anger, Chair, Cold, Cup, Rough, Sleep, Smell, Smoke, Soft, Sweet, Trash and Window. According to the norms from Nelson, McEvoy & Schreiber (2004), the mean Backward Associative Strength (BAS) to the critical lures for the included studied list items was .197 (*SD* = .097, norms available for 6-11 items per list).

An additional 30 unrelated words were drawn from the MRC Psycholinguistic Database (M. D. Wilson, 1988) to serve as unrelated novel items. To reproduce Tun et al.'s (1998) procedure as closely as possible we matched the novel items to the list items but did not counterbalance their study-test status over participants, although as explained in the Introduction (see also (Burnside et al., 2017) this procedure does not offer full control of baseline false recognition. The unrelated words were matched to the list items on Kucera-Francis written frequency (for list items *M* = 72 per million, for unrelated lures *M* = 69), familiarity (for list items *M* = 551, for unrelated lures *M* = 547) and imageability (for list items *M* = 527, for unrelated lures *M* = 516). The words were recorded by a female native speaker of English.

Each study list included the top 12 associates for one critical lure. Each recognition test list consisted of half studied (‘old’) list items and half unstudied items. The Studied items were from list positions 2, 4, 5, 6, 7, 9 and 10. The unstudied items included the Critical lure for that list, 3 Weak lures from list positions 13, 14 and 15, and 3 Unrelated words taken from the matched unrelated item set. The test items were placed in a different random order for each of 6 stimulus sets. The sets were randomly allocated to participants.

Procedure

Participants were tested individually seated at a computer and first completed the WTAR and demographic questions, and then the experimental task. Prior to the experiment, the sound level of the stimuli was adjusted to be comfortable for each participant. All wore headphones except one who preferred not to. Participants were instructed not to respond to any items in the memory test which they could not hear clearly.

The experiment comprised 10 study-test cycles. Participants were instructed to remember as many words from each study list as possible. Lists were presented auditorily with words 1.5 sec apart. After each list, participants completed the corresponding recognition test. At test, words were again presented auditorily but the task was self-paced, with a response-to-stimulus interval of 2 sec. Participants were asked to judge whether each item had been presented on the preceding study list, and to indicate their responses using key presses, prioritizing accuracy. Study lists were always presented in the order: Window, Smell, Cold, Rough, Cup, Soft, Sleep, Anger, Sweet, Trash, Chair, Smoke. Response hands were counterbalanced across participants. Stimulus presentation was programmed using Cogent2000 (Romaya, 2000) (version 1.32), and MATLAB (R2015b, www.themathworks.com).

Data analysis

Pre-processing. Participants were excluded from analysis according to the following *a priori* criteria (see Participants): if the mean hit proportion across conditions was less than 10% greater than the mean proportion of false alarms to unrelated lures, if there were fewer than 5 valid trials in any condition (i.e., excluding trials with no responses or multiple different responses), or if they were an outlier (> 3 SD from the mean) on any d' measure.

Statistical analysis. Except where stated, analytic procedures were the same as in Experiment 1. To follow the original procedures of Tun et al.'s (1998) Experiment 2, the main outcome measures were raw response proportions and mean response times (RTs) for items attracting 'old' responses (items judged as studied). RTs were computed where there were at least 5 trials per condition. We also conducted converging analyses using d' measures, as in Experiment 1, and an exploratory analysis of *z*-score transformed RTs, which is reported in the Supplemental Material.

Results

Preregistered analysis

Table 2 lists summary data for the raw proportions of "old" responses and response times in Experiment 2.

[Table 2 near here]

True and false recognition: raw accuracy. Analysis of raw response proportions did not reveal any significant effects of age. ANOVA with factors of Age and Item type (Hit, Critical lure FA, Weak lure FA, Unrelated FA) gave rise only to a main effect of Item type ($F(2.1,127.0) = 950$, $p < 0.001$, $\eta^2_p = .940$; for main effect of Age, $F(1,61) = .743$, $p = 0.392$; for interaction, $F(2.1,127.0) = .317$, $p = 0.737$). *Post hoc* tests showed significant differences in the proportions of ‘old’ responses between all pairs of item types: for Hits and Critical FA, $t(61) = 4.56$, Bonferroni-corrected $p < 0.001$; for Critical FA vs. Weak FA, $t(61) = 27.1$, $p < 0.001$, for Weak FA vs. Unrelated FA, $t(61) = 7.85$, $p < 0.001$). This suggests that participants across both age groups responded ‘old’ significantly more often to Studied items than to any unstudied items, including Critical lures. There was also strong Bayesian evidence against an interaction of Age and Item type ($BF_{01} = 18.1$).

Since Tun et al.’s (1998) main claim was of a specific age-related difference in Critical FA, we also conducted a separate test of the interaction of Age x Item type in an ANOVA including only Hits and Critical FA. This, too, yielded a null effect ($F(1,61) = .100$, $p = 0.753$) and moderate evidence favouring the null ($BF_{01} = 3.41$).

True and false recognition: response times. We did not decide *a priori* on a minimum number of participants per group for the RT analysis, but followed the original study in comparing only studied items and critical lures, for which 27 young and 32 older participants had enough trials for each condition. ANOVA with factors of Age and Item type (Hit, Critical lure FA) revealed slower responses to Critical lures than to studied items but this effect was indistinguishable between the two age groups (for main effect of Item type, $F(1,57) = 18.23$, $p < 0.001$, $\eta^2_p = .246$; for interaction with Age $F(1,57) = .110$, $p = 0.741$). Surprisingly, the older adults did not respond significantly more slowly overall in this task (for main effect of Age, $F(1,57) = .144$, $p = 0.706$). There was moderate Bayesian evidence against an interaction of Age and Item Type ($BF_{01} = 3.98$; for z-transformed RTs, $BF_{01} = 5.21$; see Supplemental Material).

True and false recognition: discrimination. Analysis of true and false lure discrimination using d’ yielded a similar picture to the analysis of raw false recognition. Figure 2 illustrates the results for critical lures. ANOVA with factors of Age and Item type (True d’, Critical lure d’, Weak lure d’) gave rise only to a main effect of Item type ($F(1.98,120.9) = 597$, $p < 0.001$,

η^2_p = .907; for main effect of Age, $F(1,61) = 1.03$, $p = 0.315$; for interaction, $F(1.98,120.9) = .167$, $p = 0.844$). *Post hoc* tests showed that true memory was better than false memory even though raw proportions of hits and Critical FA did not differ (see above; for True d' vs. Critical lure d', $t(61) = 6.84$, Bonferroni-corrected $p < 0.001$; for Critical lure d' vs. Weak lure d', $t(61) = 25.1$, $p < 0.001$). There was strong Bayesian evidence against an interaction of Age x Item type ($BF_{01} = 31.5$). A separate comparison for Critical lure d' also showed moderate evidence against an age-related difference ($BF_{01} = 5.20$; ; with prior width of .5, $BF_{01} = 3.92$).

[Figure 2 near here]

Exploratory analysis

Discrimination between studied items and lures. As in Experiment 1, we also tested whether discrimination of lures relative to studied items was reduced in older compared to young adults, using a corresponding d' measure. There was anecdotal evidence against such an effect ($BF_{01} = 2.45$).

Discussion

Experiment 2 used a typical DRM procedure with auditory presentation and intentional encoding with short SOA, as well as blocked word lists. Aside from restricting the memory test to recognition, we based the materials and procedure as closely as possible on Tun et al.'s (1998) Experiment 2. The current study had substantially larger samples than the original, so we had good power to detect a moderate-sized difference between groups, i.e., an effect more than three times smaller than that originally reported (see Participants). Tun et al. (1998, Experiment 2) found that older adults made more false alarms to critical lures than the young, with no significant difference for unrelated novel items (although these two item types were not directly compared). The slowing of responses to critical lures relative to studied items was also greater in the original older group. We found evidence against both these effects regardless of whether overall age effects on RTs were removed by a z-transformation, unlike in Experiment 1 in which evidence against age effects on RTs was only clear-cut for transformed data. The findings also converged across both raw false recognition proportions (analysed as in Tun et al.'s (1998) study) and discrimination measures which properly corrected for response bias. The data therefore suggest that associative false recognition does not increase with age even when a typical DRM procedure likely to emphasise age-related differences is used.

General Discussion

These two studies challenge the widely held view that older adults rely more on semantic gist in memory than young adults (e.g., Brainerd, Reyna, & Howe, 2009; Devitt & Schacter, 2016; Schacter et al., 1997; Tun et al., 1998). False recognition was equivalent in young and older adults even under conditions held to favour gist as opposed to item-specific processing. Experiment 1 showed age-invariance under blocked presentation in a visual task in which multiple semantically associated lists were studied with an orienting task which encourages item-specific over relational encoding (Smith & Hunt, 1998). This suggests that list blocking by itself is not critical (see also Coane et al., 2016). Moreover, false recognition remained age-invariant in Experiment 2, which followed Tun et al.'s (1998, Experiment 2) typical DRM procedure with intentional encoding of auditory words, conditions that should maximise age differences. Together, the results suggest that older people are not more prone to associative false recognition.

Our findings do not of course rule out age effects on associative false recall: there is relatively consistent evidence for these (but see e.g., Coane et al., 2016). In Gallo's (2006) analysis, older people falsely recalled 7% more critical lures and made 15% more other intrusions than young people. Free recall, unlike simple recognition, is thought to place high demands on pre-retrieval control functions such as initial cue elaboration and search, since there is no specific external retrieval cue. Post-retrieval source monitoring is also more taxing because the recalled information being evaluated in relation to the study episode is less detailed than an externally presented recognition test probe (Johnson, Hashtroudi, & Lindsay, 1993). Both pre- and post-retrieval control processes are thought to be impaired in ageing (Craik, 1983; Gallo, 2006; Morcom, 2016): for example, older adults are less likely to engage post-retrieval strategies like disqualifying recall-to-reject monitoring (Dehon & Brédart, 2004; Gallo, Bell, Beier, & Schacter, 2006). Thus, age-related recall errors are thought to reflect impaired source monitoring rather than enhanced activation of semantic associations or gist. These findings are consistent with the theory that age effects on memory are secondary to declining frontal lobe function (West, 1996), leading to impaired cognitive control with no concomitant changes in representation. However, as noted in the Introduction, there is converging evidence that ageing does also affect mental representations, which may explain findings that older adults show heightened false recognition of other types of recognition lures. An important caveat is that the current data do not speak directly to the question of whether prior recall inflates false recognition, and whether it does so to a greater degree in older adults (see Introduction). However, this is distinct from wider questions concerning age effects on false recognition.

Together the data support the suggestion in the Introduction that there is a contrast between the age effects found in the DRM and categorised pictures paradigms. Unlike in the DRM task, older people are consistently more likely to falsely recognise categorically related picture lures (typically different exemplars of a basic level category), even though intermixed designs and item-specific encoding are used (Koutstaal et al., 2003; Koutstaal & Schacter, 1997; Pidgeon & Morcom, 2014). There is also evidence for a critical role of semantic relations in the categorized pictures task (see Introduction; Koutstaal et al., 2003; Pidgeon & Morcom, 2014). However, the current findings challenge the more general proposal that older adults rely more on prior knowledge in memory tasks than the young (Castel, 2005; Naveh-Benjamin et al., 2005; Reder et al., 2007; Umanath & Marsh, 2014). This more nuanced view converges with results from studies using other kinds of memory task that suggest that semantic knowledge interacts with episodic memory in multiple ways, not all of which are impacted by ageing (Badham, Estes, & Maylor, 2012; Badham, Hay, Foxon, Kaur, & Maylor, 2016). In the current study, discrimination of studied items from unrelated lures (as well as true recognition) was also age-invariant, which weighs against the generality of the recent meta-analytic finding that age effects on discrimination of studied items from lures are larger when lures are semantically-related than unrelated (Fraundorf et al., 2019). One source of differences in findings may be that some studies included in the meta-analysis tested recall prior to recognition, rather than assessing recognition alone.

One potential explanation for the apparent differences between the two main paradigms used to study false recognition in aging is that the specific nature of the semantic relations between lures and studied items is critical for whether age effects are observed, and older adults may make more errors only when lures and studied items are similar to one another. As noted earlier, while DRM lures are semantically related to items on the study lists as well as being strongly associated (Brainerd et al., 2008; Cann et al., 2011), they are often dissimilar, unlike items in categorised pictures tasks which share many features (e.g., different dogs may all have legs, a tail, and fur, and be able to bark). In young adults, shared semantic features appear to make a contribution to false recognition that is separable from the effects of associations, at least for verbal material (Coane et al., 2015; Montefinese et al., 2014). If similarity is critical for age effects, older adults should show increased false recognition of categorised word stimuli, just as they do for pictures. It is currently unclear whether this is the case (although Huff & Aschenbrenner, 2018, and Dennis, Kim, & Cabeza, 2007) both found non-significant age-related differences). Direct comparison of age effects according to lure relations using the same stimuli will be critical in testing this proposal.

If age effects on false recognition are related to similarity between lures and studied items, one possibility is that demands on post-retrieval monitoring are greater when similarity is high (Coane et al., 2015). This suggestion has not directly been tested, but recent data from Trelle et al. (2018) point to an age effect on false recognition of categorically related pictures even when monitoring demands are minimised using a forced choice recognition test. Other potential explanations draw on theories proposing aging effects on mental representation as well as cognitive control (Shing, Werkle-bergner, Li, & Lindenberger, 2008). Reduced specificity of memory representations with age may reflect impaired hippocampal pattern separation of memories for experiences that are similar along one or more dimensions (I. A. Wilson et al., 2006; Yassa, Lacy, Stark, Albert, & Stark, 2011). Alternatively, there may be a pervasive reduction in the specificity of representations for perceptual as well as memory representations (dedifferentiation; Li et al., 2001), although evidence for this is not clear cut (e.g., Abdulrahman et al., 2014; Trelle et al., 2017). Either can account for findings that older people are less able to reject lures which are the same objects spatially displaced from an original location (Reagh et al., 2014), or separated in time (Roberts, Ly, Murray, & Yassa, 2014), as well as when lures are similar objects (Yassa et al., 2011). A more specific proposal is that of a loss of integrity in cross-modal conjunctive feature representations that support fine-grained distinctions between objects (Burke & Barnes, 2018). These views all predict that older adults will be more prone to falsely recognise recognition lures that are perceptually as well as conceptually similar, for verbal as well as pictorial material (Burnside et al., 2017; Pidgeon & Morcom, 2014; Trelle et al., 2017; D. M. Wilson et al., 2018).

The divergence between age effects on false recognition between DRM and categorised pictures tasks and materials is more in line with Fuzzy Trace Theory than Activation Monitoring Theory. Several studies point to effects of conceptual relations on memory over and above effects of associative relations, suggesting that an associative activation account is insufficient (Cann et al., 2011; Carneiro et al., 2017; Coane et al., 2015; Montefinese et al., 2014). Fuzzy Trace theory also explicitly accommodates contributions of multiple different semantic relations to gist effects in memory, in the sense that the relevant gist shared by studied items and lures depends on the stimuli used and how they are related: for example, while some may share a category, others share a situational theme (Cann et al., 2011), or a narrative (Reyna, Corbin, Weldon, & Brainerd, 2016). More detailed specification of these dimensions of gist may help explain why a variable like aging might impact one and not another. For example, dissociations of memory effects due to similarity and association would be consistent with the more fundamental proposal that semantic memory is based on both categorical and thematic relations, which are represented differently in the brain (Carneiro et al., 2017; Mirman, Landrigan, & Britt, 2017). Future studies should address these possibilities directly.

In summary, there is no doubt that older people do make more memory errors than young people in a number of tasks. But the two studies reported here support the proposal that associative false recognition is an important exception.

Data availability statement

All data and experimental materials are publicly available on the Open Science Foundation site for both experiments (for Experiment 1, DOI 10.17605/OSF.IO/R5NQU; for Experiment 2, DOI 10.17605/OSF.IO/VXAEH).

References

Abdulrahman, H., Fletcher, P. C., Bullmore, E., & Morcom, A. M. (2014). Dopamine and memory dedifferentiation in aging. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2015.03.031

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2018). Raincloud plots:

a multi-platform tool for robust data visualization. *PeerJ Preprints*, *6*(e27137v1).

Arndt, J. (2010). The role of memory activation in creating false memories of encoding context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 66–79. https://doi.org/10.1037/a0017394

Arndt, J., & Reder, L. M. (2003). The effect of distinctive visual information on false recognition. *Journal of Memory and Language*, *48*(1), 1–15. https://doi.org/10.1016/S0749-596X(02)00518-1

Badham, S. P., Estes, Z., & Maylor, E. A. (2012). Integrative and semantic relations equally alleviate age-related associative memory deficits. *Psychology and Aging*, *27*(1), 141–152. https://doi.org/10.1037/a0023924

Badham, S. P., Hay, M., Foxon, N., Kaur, K., & Maylor, E. A. (2016). When does prior knowledge disproportionately benefit older adults memory? *Aging, Neuropsychology, and Cognition*, *23*(3), 338–365. https://doi.org/10.1080/13825585.2015.1099607

Brainerd, C. J., Reyna, V. F., & Howe, M. L. (2009). Trichotomous processes in early memory development, aging, and neurocognitive impairment: a unified theory. *Psychological Review*, *116*(4), 783–832. https://doi.org/10.1037/a0016963

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychonomic Bulletin & Review*, *15*(6), 1035–1053. https://doi.org/10.3758/PBR.15.6.1035

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer’s disease. *Brain and Cognition*, *51*(3), 251–261. https://doi.org/10.1016/S0278-2626(03)00030-7

Burke, S. N., & Barnes, C. A. (2018). *The Contribution of Recollection, Familiarity and Discrimination to Object Recognition Deficits in Advanced Age*. *Handbook of Behavioral Neuroscience* (Vol. 27). Elsevier. https://doi.org/10.1016/B978-0-12-812012-5.00020-3

Burnside, K., Hope, C., Gill, E., & Morcom, A. M. (2017). Effects of perceptual similarity but not semantic association on false recognition in aging. *PeerJ*, *2017*(12). https://doi.org/10.7717/peerj.4184

Cann, D. R., McRae, K., & Katz, A. N. (2011). False recall in the Deese-Roediger-McDermott paradigm: The roles of gist and associative strength. *Quarterly Journal of Experimental Psychology (2006)*, *64*(8), 1515–1542. https://doi.org/10.1080/17470218.2011.560272

Carneiro, P., Garcia-Marques, L., Lapa, A., & Fernandez, A. (2017). Explaining the persistence of false memories: a proposal based on associative activation and thematic extraction. *Memory*, *25*(8), 986–998. https://doi.org/10.1080/09658211.2016.1239742

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: the role of schematic support. *Psychology and Aging*, *20*(4), 718–721. https://doi.org/10.1037/0882-7974.20.4.718

Coane, J. H., Huff, M. J., & Hutchison, K. A. (2016). The ironic effect of guessing: Increased false memory for mediated lists in younger and older adults. *Aging, Neuropsychology, and Cognition*, *23*(3), 282–303. https://doi.org/10.1080/13825585.2015.1088506

Coane, J. H., McBride, D. M., Termonen, M.-L., & Cutting, J. C. (2015). Categorical and associative relations increase false memory relative to purely associative relations. *Memory & Cognition*, *43*(6), 1–13. https://doi.org/10.3758/s13421-015-0543-1

Craik, F. I.M., & Rabinowitz, J. C. (1985). The effects of presentation rate and encoding task on age-related memory deficits. *Journals of Gerontology*, *40*(3), 309–315.

https://doi.org/10.1093/geronj/40.3.309

Craik, Fergus I M. (1983). On the Transfer of Information from Temporary to Permanent Memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *302*(1110), 341–359. https://doi.org/10.1098/rstb.1983.0059

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *5*, 305–312. https://doi.org/10.2466/pr0.1959.5.3.305

Dehon, H., & Brédart, S. (2004). False Memories: Young and Older Adults Think of Semantic Associates at the Same Rate, but Young Adults Are More Successful at Source Monitoring. *Psychology and Aging*, *19*(1), 191–197. https://doi.org/10.1037/0882-7974.19.1.191

Dennis, N. A., Kim, H., & Cabeza, R. (2007). Effects of aging on true and false memory formation : An fMRI study. *Neuropsychologia*, 45(14), 3157-3166. https://doi.org/10.1016/j.neuropsychologia.2007.07.003

Devitt, A. L., & Schacter, D. L. (2016). False memories with age_ neural and cognitive underpinnings. *Neuropsyhology*, *91*, 346–359. https://doi.org/10.1016/j.neuropsychologia.2016.08.030

Faul, F., Erdfelder, E., Lang, A.-G. G., Buchner, A., & Kiel, C. (2007). G*Power3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences. *Behavioral Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/bf03193146

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*.

Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2019). Aging and recognition memory: A meta-analysis. *Psychological Bulletin*, *145*(4), 339–371. https://doi.org/10.1037/bul0000185

Gallo, D. A. (2006). *Associative Illusions of Memory*. New York: Psychology Press.

Gallo, D., Bell, D., Beier, J., & Schacter, D. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory, 14*(6), 730–741. https://doi.org/10.1080/09658210600648506

Gardiner, J. M., & Java, R. I. (1991). Forgetting in recognition memory with and without recollective experience. *Memory and Cognition*, 19(6), 617-623.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Holdnack, H. A. (2001). *Wechsler Test of Adult Reading: WTAR*. San Antonio: The Psychological Corporation.

Huff, M. J., & Aschenbrenner, A. J. (2018). Item-specific processing reduces false recognition in older and younger adults: Separating encoding and retrieval using signal detection and the diffusion model. *Memory and Cognition*, *46*(8), 1287–1301. https://doi.org/10.3758/s13421-018-0837-1

Huff, M. J., Coane, J. H., Hutchison, K. A., Grasser, E. B., & Blais, J. E. (2012). Interpolated task effects on direct and mediated false recognition: Effects of initial recall, recognition, and the ironic effect of guessing. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(6), 1720–1730. https://doi.org/10.1037/a0028476

Huff, M. J., & Hutchison, K. A. (2011). The effects of mediated word lists on false recall and recognition. *Memory and Cognition*, *39*(6), 941–953. https://doi.org/10.3758/s13421-011-0077-0

Hutchison, K. A., & Balota, D. A. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous and unambiguous critical lures. *Journal of Memory and Language*, *52*(1), 1–28. https://doi.org/10.1016/j.jml.2004.08.003

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, a. (1999). Age, testing at preferred or nonpreferred times (testing optimality), and false memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(1), 23–40. https://doi.org/10.1037/0278-7393.25.1.23

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source Monitoring. *Psychological Bulletin*, *114*(1), 3-28.

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: effects of ageing. *Cognitive Neuropsychology*, *16*(3–5), 399–415.

Koutstaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: testing the semantic categorization account. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*(4), 499–510. https://doi.org/10.1037/0278-7393.29.4.499

Koutstaal, W., & Schacter, D. L. (1997). Gist-Based False Recognition of Pictures in Older and Younger Adults, *Journal of Memory and Language*, 37, 555–583.

Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. a. (1999). Reducing gist-based false recognition in older adults: encoding and retrieval manipulations. *Psychology and Aging*, *14*(2), 220–237. https://doi.org/10.1037/0882-7974.14.2.220

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course.* Cambridge University Press.

Li, S.-C., Li, S.-C., Lindenberger, U., Lindenberger, U., Sikstrom, S., & Sikstrom, S. (2001). Aging cognittion: From neuromodulation to representation, *5*(11), 479–486.

McCabe, D. P., Roediger, H. L., Mcdaniel, M. A., & Balota, D. A. (2009). Aging reduces veridical remembering but increases false remembering : Neuropsychological test correlates of remember – know judgments. *Neuropsychologia*, *47*, 2164–2173. https://doi.org/10.1016/j.neuropsychologia.2008.11.025

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory and Cognition*, *30*(7), 1065–1077. https://doi.org/10.3758/bf03194324

Migo, E. M., Mayes, A. R., & Montaldi, D. (2012). Measuring recollection and familiarity: Improving the remember/know procedure. *Consciousness and Cognition*, *21*(3), 1435–1455. https://doi.org/10.1016/j.concog.2012.04.014

Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499–520. https://doi.org/10.1037/bul0000092

Montefinese, M., Zannino, G. D., & Ambrosini, E. (2014). Semantic similarity between old and new items produces false alarms in recognition memory. *Psychological Research*, 785–794. https://doi.org/10.1007/s00426-014-0615-z

Morcom, A. M. (2016). Mind Over Memory: Cuing the Aging Brain. *Current Directions in Psychological Science*, *25*(3). https://doi.org/10.1177/0963721416645536

Naveh-Benjamin, M., Craik, F. I. M., & Guez, J. (2005). Divided Attention in Younger and Older Adults : Effects of Strategy and Relatedness on Memory Performance and Secondary Task Costs, *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(3), 520–537. https://doi.org/10.1037/0278-7393.31.3.520

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: exploring the characteristics of illusory memories. *Memory and Cognition*, *25*(6), 838–848. https://doi.org/10.3758/bf03211328

Old, S. R., & Naveh-benjamin, M. (2008). Differential effects of age on item and associative measures of memory: a meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Pidgeon, L. M., & Morcom, A. M. (2014). Age-related increases in false recognition: The role of perceptual and conceptual similarity. *Frontiers in Aging Neuroscience*, *6*(OCT). https://doi.org/10.3389/fnagi.2014.00283

Reagh, Z. M., Noche, J. A., Tustison, N. J., Delisle, D., Murray, E. A., & Yassa, M. A. (2018). Functional Imbalance of Anterolateral Entorhinal Cortex and Hippocampal Dentate/CA3 Underlies Age-Related Object Pattern Separation Deficits. *Neuron*, *97*(5), 1187-1198.e4. https://doi.org/10.1016/j.neuron.2018.01.039

Reagh, Z. M., Roberts, J. M., Ly, M., Diprospero, N., Murray, E., & Yassa, M. A. (2014). Spatial discrimination deficits as a function of mnemonic interference in aged adults with and without memory impairment. *Hippocampus*, *24*(3), 303–314.

https://doi.org/10.1002/hipo.22224

Reder, L. M., Paynter, C. A., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a double-edged sword: a computational model of the encoding/ retrieval trade-off with familiarity. In B. Rossand & A. Benjamin (Eds.), *The Psychology of Learning and Motivation* (pp. 271–312). Academic Press.

Reyna, V F, & Brainerd, J. (1995). Fuzzy Trace Theory: An interim synthesis. *Learning and Individual Differences*, *7*(1), 1–75.

Reyna, Valerie F., Corbin, J. C., Weldon, R. B., & Brainerd, C. J. (2016). How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. *Journal of Applied Research in Memory and Cognition*, *5*(1), 1–9. https://doi.org/10.1016/j.jarmac.2015.12.003

Roberts, J. M., Ly, M., Murray, E., & Yassa, M. A. (2014). Temporal discrimination deficits as a function of lag interference in older adults. *Hippocampus*, *24*(10), 1189–1196. https://doi.org/10.1002/hipo.22303

Roediger, H. L., & McDermott, K. B. (1995). Creating False Memories: Remembering Words Not Presented in Lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814.

Roediger III, H L, Watson, J. M., McDermott, K. B., Gallo, D. a, Roediger, H. L., Watson, J. M., … Gallo, D. a. (2001). Factors that determine false recall: a multiple regression analysis. *Psychonomic Bulletin and Review*, *8*(3), 385–407. https://doi.org/10.3758/BF03196177

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing False Recognition in Younger and Older Adults: The Distinctiveness Heuristic. *Journal of Memory and Language*, *40*(1), 1–24. https://doi.org/10.1006/jmla.1998.2611

Schacter, D. L., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences*, *6613*(6), 229–236.

Shing, Y. L., Werkle-bergner, M., Li, S., & Lindenberger, U. (2008). Associative and Strategic Components of Episodic Memory : A Life-Span Dissociation, *137*(3), 495–513. https://doi.org/10.1037/0096-3445.137.3.495

Skinner, E. I., & Fernandes, M. A. (2009). Illusory Recollection in Older Adults and Younger Adults Under Divided Attention. *Psychology and Aging*, *24*(1), 211–216. https://doi.org/10.1037/a0014177

Smith, R. E., & Hunt, R. R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, *5*, 710–715. https://doi.org/10.3758/BF03208850

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of Measuring Recognition Memory : Applications to Dementia and Amnesia, (I), 34–50.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: a meta-analysis. *Psychology and Aging*, *10*(4), 527–539. https://doi.org/10.1037/0882-7974.10.4.527

Stark, C. E. L., & Squire, L. R. (2000). Functional Magnetic Resonance Imaging (fMRI) Activity in the Hippocampal Region during Recognition Memory, *20*(20), 7776–7781.

Taconnat, L., Clarys, D., Vanneste, S., & Isingrini, M. (2006). Effect of distinctive encoding on false recognition, discrimination, and decision criteria in young and elderly adults. *European Journal of Cognitive Psychology*, *18*(5), 708–729. https://doi.org/10.1080/09541440500285585

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52*(1), 71–86. https://doi.org/10.1016/j.jml.2004.08.001

Trelle, A. N., Henson, R. N., Green, D. A. E., & Simons, J. S. (2017). Declines in Representational Quality and Strategic Retrieval Processes Contribute to Age-Related Increases in False Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000412

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging*, *13*(2), 230–241. https://doi.org/10.1037/0882-7974.13.2.230

Umanath, S., & Marsh, E. J. (2014). Understanding How Prior Knowledge Influences Memory in Older Adults. *Perspectives on Psychological Science*, *9*(4), 408–426. https://doi.org/10.1177/1745691614535933

Vega-Mendoza, M., West, H., Sorace, A., & Bak, T. H. (2015). The impact of late, non-balanced bilingualism on cognitive performance. *Cognition*, *137*, 40–46. https://doi.org/10.1016/j.cognition.2014.12.008

Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, *15*(2), 254–267. https://doi.org/10.1037/0894-4105.15.2.254

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272–292. https://doi.org/10.1037/0033-2909.120.2.272

Wilson, D. M., Potter, K. W., & Cowell, R. A. (2018). Recognition memory shielded from semantic but not perceptual interference in normal aging. *Neuropsychologia*, *119*(July), 448–463. https://doi.org/10.1016/j.neuropsychologia.2018.07.031

Wilson, I. A., Gallagher, M., Eichenbaum, H., & Tanila, H. (2006). Neurocognitive aging: prior memories hinder new hippocampal encoding. *Trends in Neurosciences*, *29*(12), 662–670. https://doi.org/10.1016/j.tins.2006.10.002

Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, *20*, 6–11.

Wong, J. T., & Gallo, D. A. (2016). Stereotype threat reduces false recognition when older adults are forewarned. *Memory*, *24*(5), 650–658. https://doi.org/10.1080/09658211.2015.1036885

Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. L. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(21), 8873–8878. https://doi.org/10.1073/pnas.1101567108

Yassa, M. a, Lacy, J. W., Stark, S. M., Albert, M. S., & Stark, C. E. L. (2011). Pattern Separation deficits associated with increased hippocampal CA3 and dentate gyrus activity in nondemented older adults. *Hippocampus*, *21*(9), 968–979. https://doi.org/10.1002/hipo.20808.Pattern

Tables

Table 1

Accuracy and response times for Experiment 1. Means and standard deviations (in brackets) are listed for accuracy proportions and response times (RTs) by age group, condition and trial type.

	DRM		DRM-font		Font	
	Young	Older	Young	Older	Young	Older
	M *(SD)*	M *(SD)*	M *(SD)*	M *(SD)*	M *(SD)*	M *(SD)*
N	34	33	31	29	38	23
Proportion "old" judgments						
Studied items	0.936	0.906	0.960	0.939	0.934	0.933
	(0.077)	(0.068)	(0.029)	(0.053)	(0.060)	(0.057)
Critical lures	0.524	0.426	0.567	0.531	0.225	0.164
	(0.236)	(0.244)	(0.263)	(0.295)	(0.196)	(0.145)
Weak lures	0.240	0.160	0.288	0.304	0.188	0.126
	(0.140)	(0.138)	(0.228)	(0.215)	(0.163)	(0.095)
Novel (critical-matched)	0.118	0.086	0.099	0.102	0.116	0.095
	(0.095)	(0.105)	(0.079)	(0.167)	(0.109)	(0.080)
Novel (weak-matched)	0.141	0.107	0.095	0.125	0.132	0.100
	(0.113)	(0.092)	(0.061)	(0.127)	(0.107)	(0.073)
Discrimination (d')						

True vs. novel	2.88	2.78	3.22	2.93	2.87	2.99
	(0.676)	(0.606)	(0.462)	(0.710)	(0.743)	(0.551)
Critical vs. novel	1.37	1.23	1.59	1.52	0.399	0.292
	(0.794)	(0.828)	(0.826)	(0.821)	(0.619)	(0.561)
Weak vs. novel	0.396	0.213	0.703	0.664	0.200	0.121
	(0.599)	(0.505)	(0.822)	(0.513)	(0.733)	(0.354)
True vs. critical	1.79	1.82	1.85	1.73	2.68	2.54
	(0.736)	(1.13)	(0.763)	(0.915)	(0.770)	(0.776)
RT for "old" judgments						
Studied items	1220	1700	1140	1680	1210	1910
	(462)	(450)	(264)	(413)	(361)	(511)
Critical lures	2490	4070	2310	2590	2730	3280
	(1290)	(2220)	(1240)	(899)	(1410)	(1943)

Note. *M* = mean; *SD* = standard deviation. For calculation of d' see Methods: Data analysis.

Table 2

Accuracy and response times for Experiment 2. Means and standard deviations (in brackets) are listed for accuracy proportions and response times (RTs) by age group and trial type.

	Young M *(SD)*	Older M *(SD)*
Proportion "old" judgments		
Studied items	0.866 (0.080)	0.831 (0.090)
Critical lures	0.750 (0.217)	0.741 (0.155)
Weak lures	0.151 (0.141)	0.127 (0.115)
Novel items	0.031 (0.042)	0.030 (0.039)
Discrimination (d')		
True recognition	2.98 (0.496)	2.83 (0.519)
Critical lure false recognition	2.47 (0.641)	2.40 (0.572)
Weak lure false recognition	0.743 (0.485)	0.620 (0.499)
True versus critical lure discrimination	0.500 (0.645)	0.424 (0.410)
RT for "old" judgments		
Studied items	1270 (197)	1300 (267)
Critical lures	1560 (564)	1550 (477)

Figures

Figure 1

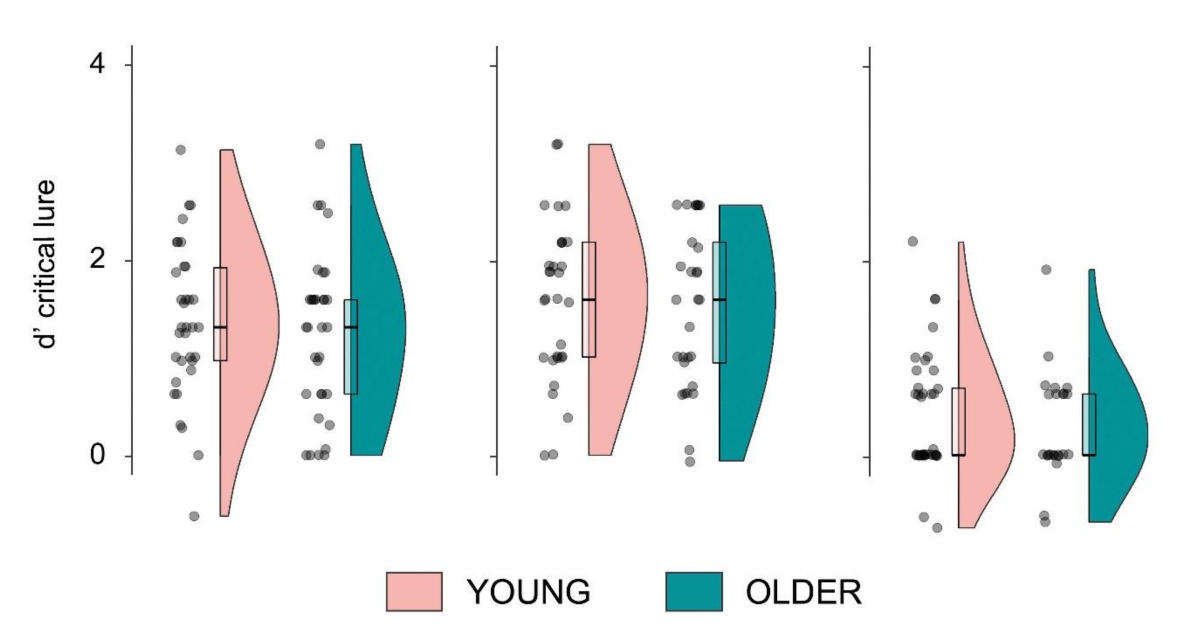

Figure 2

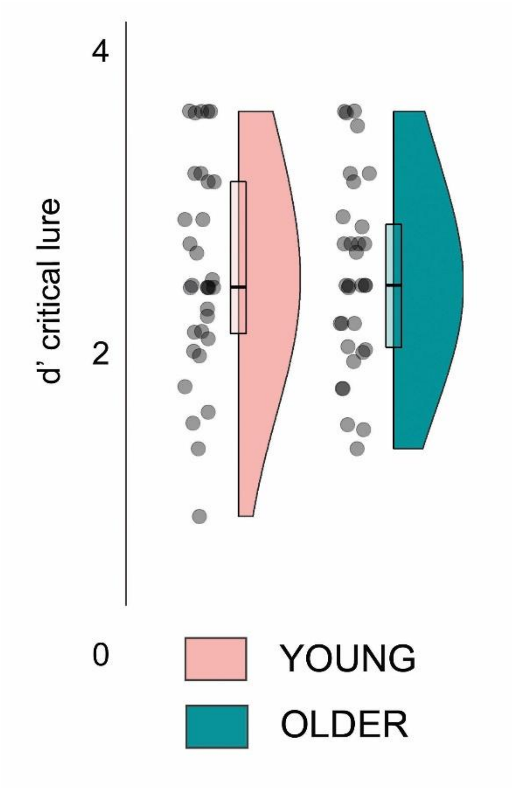

Figure captions

Figure 1. False recognition of critical lures in Experiment 1 by young and older adults. Panels show discrimination (d') of critical lures from matched novel items in the DRM-only (left), DRM-font (center) and Font-only (right) conditions (see Data analysis for measures). Within each panel, filled circles show raw data with horizontal jitter, shaded areas show the probability density function of the data, and boxplots show the median and interquartile range.

Figure 2. False recognition of critical lures in Experiment 2 by young and older adults. Plots show discrimination (d') of critical lures from matched novel items (see Data analysis). Filled circles show raw data with horizontal jitter, shaded areas show the probability density function of the data, and boxplots show the median and interquartile range.